

# ON THE DISCRETE LOGARITHM PROBLEM

CRISTIAN COBELI

**ABSTRACT.** Let  $p > 2$  be prime and  $g$  a primitive root modulo  $p$ . We present an argument for the fact that discrete logarithms of the numbers in any arithmetic progression are uniformly distributed in  $[1, p]$  and raise some questions on the subject.

## 1. INTRODUCTION

Before the middle of the last century, discrete logarithms were just common tools used to perform calculations in finite fields. Then, with the development of cryptography, their importance raised considerably, especially after Diffie and Hellman [1] created the key exchange algorithm, the first practical public key cryptosystem. Many cryptosystems, such as the Diffie-Hellman key agreement and its derivatives, ElGamal public-key encryption, ElGamal signature scheme and its variants, DSA, etc. (see [2], [3], [4]) are based on the assumption that discrete logarithms are hard to compute. Considerable efforts have been made to find algorithms that speed up the calculation of discrete logarithms, but nobody knows how one could prove that a very fast algorithm does not exist.

A strong argument would require proofs for the random distribution characteristic of the set containing the discrete logarithms of the elements of a “regular” subset of  $[0, p - 1]$  (a subinterval being just the first try), when  $p \rightarrow \infty$ . This feature is suggested by numerical evidences for small  $p$  and by most of the work done around the cryptosystems based on discrete logarithms (see [4] and [6] and the references within). Recently, Banks and Shparlinski [7] have obtained nice results in this direction.

Discrete logarithms can be defined in general groups, but we reduce here only to the group  $\mathcal{G} = \mathbb{Z}/p\mathbb{Z}$  of residue classes modulo a prime  $p > 2$ . Given any  $g \in \mathcal{G}$  and  $n \in \mathbb{N}$ , let  $g^n := \underbrace{g \cdots g}_n$  be the discrete exponentiation function. We will assume that  $g$  is a generator

---

2000 *Mathematics Subject Classification.* Primary 11A07; Secondary 11B50, 11L07 .

Key Words and Phrases: Discrete logarithms, exponential sums, characters, primitive roots.

of  $\mathcal{G}$ , that is,  $g$  is a primitive root modulo  $p$ . Then, for any  $x \in \mathcal{G}$ , the *discrete logarithm problem* requires to find the smallest integer with the property that  $g^n \equiv x \pmod{p}$ . Since  $g$  is a primitive root, the power  $n$  always exists in the interval  $[0, p-1]$ . We denote it by  $n = \log_g x = \log x$ , and call it the *discrete logarithm* of  $x$  to base  $g$ .

Notice that the discrete logarithm function is the inverse of the discrete exponentiation function and it has the properties  $\log 1 = 0$  and  $\log xy = \log x + \log y \pmod{p-1}$ , for any  $x, y \in \mathcal{G}$ .

Let  $a \geq 0$ ,  $r > 0$ ,  $N > 0$  be integers, and set  $\mathcal{J} = \{a+r, \dots, a+Nr\} \subset [1, p-1]$ . Denote

$$\mathcal{L}(g, \mathcal{J}) = \mathcal{L}(g) := \{ \log_g(a+jr) : 1 \leq j \leq N \}$$

and

$$\mathcal{M}(g, \mathcal{J}) = \mathcal{M}(g) := \left\{ \frac{\log_g(a+jr)}{p-1} : 1 \leq j \leq N \right\},$$

the image of  $\mathcal{J}$  in the torus  $\mathbb{R}/\mathbb{Z}$ . Then, any property regarding the spreading of the elements of  $\mathcal{L}(g, \mathcal{J})$  over  $[0, p-1]$  transfers into a similar one regarding the elements of  $\mathcal{M}(g, \mathcal{J})$  over the torus, and conversely. Since our aim is to understand what happens when  $p$  gets large, and it is more convenient to work within the bounded space  $\mathbb{R}/\mathbb{Z}$ , in the following our focus will concentrate mainly on  $\mathcal{M}(g, \mathcal{J})$ .

The *discrepancy* of  $\mathcal{M}(g)$  is defined by

$$\mathcal{D}(\mathcal{M}(g); \alpha, \beta) := \text{card}(\mathcal{M}(g) \cap [\alpha, \beta]) - (\beta - \alpha) \text{card}(\mathcal{M}(g)),$$

where  $0 \leq \alpha \leq \beta \leq 1$ . In order to prove that  $\mathcal{M}(g)$  is approximately uniformly distributed, which is the same as saying that  $\mathcal{J}$  is *uniformly distributed* in  $[1, p]$ , we have to show that the *extreme discrepancy*

$$\mathcal{D}(\mathcal{M}(g)) := \frac{1}{\text{card}(\mathcal{M}(g))} \sup_{1 \leq \alpha \leq \beta \leq 1} |\mathcal{D}(\mathcal{M}(g); \alpha, \beta)|$$

becomes small when  $p$  gets large. This is the object of the following theorem.

**Theorem 1.** *There exist absolute constants  $c_1, c_2 > 0$ , such that if  $\frac{1}{\pi} < p(\beta - \alpha)$ , then*

$$|\mathcal{D}(\mathcal{M}(g); \alpha, \beta)| \leq c_1 p^{1/2} \log p (2 + \log p(\beta - \alpha)) \quad (1.1)$$

for any  $0 \leq \alpha \leq \beta \leq 1$ , and

$$|\mathcal{D}(\mathcal{M}(g))| \leq \frac{c_2}{\text{card}(\mathcal{M}(g))} \cdot p^{1/2} \log^2 p. \quad (1.2)$$

A consequence of Theorem 1 assures us that any interval whose length combined with the length of  $\mathcal{J}$  exceeds a certain margin, contains plenty of elements of  $\mathcal{L}(g)$ .

**Corollary 1.** *For any  $\delta > 0$ , any subinterval of  $[0, p-1]$  of length  $M$  contains at least  $(1-\delta)\frac{MN}{p}$  and at most  $(1+\delta)\frac{MN}{p}$  elements of  $\mathcal{L}(g, \mathcal{J})$ , provided that  $MN > \frac{c_3}{\delta} p^{3/2} \log^2 p$  for some absolute constant  $c_3 > 0$ .*

## 2. ESTIMATE OF AN EXPONENTIAL SUM

One way to get bounds for the discrepancies is to obtain estimates for certain exponential sums (see (2.7) below), and this our first point.

Let  $\theta$  and  $\zeta$  be roots of unity of order  $p-1$  and  $p$ , respectively. We consider the twisted sum, called the Lagrangian resolvent of  $\theta$  and  $\zeta$ :

$$S(\theta, \zeta) := \zeta + \theta\zeta^g + \cdots + \theta^{p-2}\zeta^{g^{p-2}}. \quad (2.1)$$

Plainly  $S(1, 1) = p-1$  and it is known that

$$S(\theta, \zeta) \leq \sqrt{p}, \quad (2.2)$$

for all  $\theta$  and  $\zeta$  that are not both equal to 1. Let us see this for completeness. We have:

$$\begin{aligned} |S(\theta, \zeta)|^2 &= \sum_{k=0}^{p-2} \sum_{l=0}^{p-2} \theta^{k-l} \zeta^{g^k - g^l} \\ &= p-1 + \sum_{k=0}^{p-2} \sum_{\substack{l=0 \\ l \neq k}}^{p-2} \theta^{k-l} \zeta^{g^l(g^{k-l}-1)}. \end{aligned}$$

Let us see that here, for any  $l$  fixed, the differences  $k-l$  run over the set of nonzero classes mod  $(p-1)$ . Then, since the order of both  $\theta$  and  $g$  is  $p-1$ , the sums above are equal to

$$\begin{aligned} &= \sum_{t=1}^{p-2} \theta^t \sum_{l=0}^{p-2} \zeta^{g^l(g^t-1)} = \sum_{t=1}^{p-2} \theta^t \sum_{s=1}^{p-1} \zeta^s \\ &= \sum_{t=1}^{p-2} \theta^t \cdot (-1) = 1. \end{aligned}$$

and (2.2) follows.

By (2.1), we get

$$\sum_{j=0}^{p-2} \theta^{kj} \zeta^{u(g^j-z)} = \zeta^{-uz} S(\theta^k, \zeta^u). \quad (2.3)$$

Then we sum relations (2.3) over  $1 \leq u \leq p$ . Note that

$$\sum_{u=1}^p \zeta^{u(g^j-z)} = \begin{cases} p, & \text{if } g^j \equiv z \pmod{p}; \\ 0, & \text{otherwise,} \end{cases}$$

and observe that since  $0 \leq j \leq p-2$ , the condition  $g^j \equiv z \pmod{p}$  can be written as  $j = \log_g z$ . These yield

$$\theta^{k \log_g z} = \frac{1}{p} \sum_{u=1}^p \zeta^{-uz} S(\theta^k, \zeta^u). \quad (2.4)$$

Now taking  $\theta = e_{p-1}(1)$  and  $\zeta = e_p(1)$ , where  $e_q(x) := \exp\left(\frac{2\pi i x}{q}\right)$ , and summing equalities (2.4) over  $z \in \mathcal{J}$ , we obtain

$$\sum_{z \in \mathcal{J}} e_{p-1}(k \log_g z) = \frac{1}{p} \sum_{u=1}^p S(\theta^k, \zeta^u) \sum_{z \in \mathcal{J}} e_p(-uz). \quad (2.5)$$

The sum over  $z$  on the right-hand side is sharply bounded by

$$\begin{aligned} \left| \sum_{z \in \mathcal{J}} e_p(-uz) \right| &= \left| \sum_{j=1}^N e_p(u(a+jr)) \right| = \left| \sum_{j=1}^N e_p(ujr) \right| \\ &\leq \min \left( N, \frac{2}{|e_p(ur) - 1|} \right) \leq \min \left( N, \frac{1}{\left| \sin \frac{\pi ur}{p} \right|} \right) \\ &\leq \min \left( N, \left( 2 \left\| \frac{ur}{p} \right\| \right)^{-1} \right), \end{aligned} \quad (2.6)$$

where  $\|\cdot\|$  is the distance to the nearest integer. Then, using the (2.2) and (2.6) in (2.5), we conclude that

$$\begin{aligned} \left| \sum_{z \in \mathcal{J}} e_{p-1}(k \log_g z) \right| &\leq \frac{1}{p} \sum_{u=1}^p p^{1/2} \cdot \min \left( N, \left( 2 \left\| \frac{ur}{p} \right\| \right)^{-1} \right) \\ &\leq \sqrt{p} + p^{-1/2} \sum_{u=1}^{p-1} \left( 2 \left\| \frac{ur}{p} \right\| \right)^{-1} \\ &\leq \sqrt{p} + p^{-1/2} \sum_{v=1}^{\frac{p-1}{2}} \frac{p}{v} \leq \sqrt{p}(2 + \log p). \end{aligned} \quad (2.7)$$

The estimate (2.7) is slightly more general than the Pólya-Vinogradov inequality for character sums.

## 3. THE PROOF OF THEOREM 1 AND COROLLARY 1

A bound for the discrepancy can be deduced applying the Erdoős-Turán inequality [5, Chapter 1, page 8]. This says that for any  $0 \leq \alpha \leq \beta \leq 1$  and any positive integer  $K$ , we have

$$|\mathcal{D}(\mathcal{M}(g); \alpha, \beta)| \leq \frac{|\mathcal{M}(g)|}{K+1} + 2 \sum_{k=1}^K \left( \frac{1}{K+1} + \min \left( \beta - \alpha, \frac{1}{\pi k} \right) \right) \left| \sum_{x \in \mathcal{M}(g)} \exp(2\pi i k x) \right|.$$

Bounding the exponential sum by (2.7), the right-hand side is

$$\begin{aligned} &\leq \frac{|\mathcal{M}(g)|}{K+1} + 2\sqrt{p}(2 + \log p) \left( 1 + \sum_{k=1}^K \min \left( \beta - \alpha, \frac{1}{\pi k} \right) \right) \\ &\leq \frac{|\mathcal{M}(g)|}{K+1} + 2\sqrt{p}(2 + \log p) \left( 1 + \sum_{1 \leq k \leq \frac{1}{\pi(\beta-\alpha)}} (\beta - \alpha) + \sum_{\frac{1}{\pi(\beta-\alpha)} < k \leq K} \frac{1}{\pi k} \right) \\ &\leq \frac{|\mathcal{M}(g)|}{K+1} + c\sqrt{p} \log p \left( 1 + \left| \log K(\beta - \alpha) \right| \right), \end{aligned}$$

for some absolute constant  $c > 0$ . If we take  $K = p - 1$  in this estimate, we obtain (1.1).

Next, let us see that if  $\beta - \alpha \leq 1/\pi p$ , then  $\mathcal{M}(g)$  contains at most one element, therefore

$$\begin{aligned} \frac{1}{\text{card}(\mathcal{M}(g))} |\mathcal{D}(\mathcal{M}(g); \alpha, \beta)| &\leq \frac{1}{\text{card}(\mathcal{M}(g))} \left( 1 + (\beta - \alpha) \text{card}(\mathcal{M}(g)) \right) \\ &\leq \frac{2}{\text{card}(\mathcal{M}(g))}. \end{aligned} \tag{3.1}$$

When  $\beta - \alpha > 1/\pi p$ , we apply (1.1), and obtain

$$\begin{aligned} \frac{1}{\text{card}(\mathcal{M}(g))} |\mathcal{D}(\mathcal{M}(g); \alpha, \beta)| &\leq \frac{1}{p} + \frac{c\sqrt{p} \log^2 p}{\text{card}(\mathcal{M}(g))} \\ &\leq \frac{c' \sqrt{p} \log^2 p}{\text{card}(\mathcal{M}(g))}, \end{aligned} \tag{3.2}$$

for some absolute constant  $c' > 0$ . Now (1.2) follows from (3.1) and (3.2), and this concludes the proof of Theorem 1.

To prove Corollary 1, let  $\mathcal{I} = [s, t] \subset [0, p-1]$  be any subinterval of length  $t - s = M > 0$ , and let  $\delta > 0$ . Let  $\alpha = s/p$  and  $\beta = t/p$ . We may assume that  $\delta > 1/\sqrt{p}$ , since otherwise the result is trivial. Let  $\alpha, \beta \in [0, 1]$  with  $\beta - \alpha = N/p$ .

By the hypothesis, it follows that  $\sqrt{p} \log^2 p < c'' MN/p$  for some  $c'' > 0$ , and then by Theorem 1, it implies that  $|\mathcal{D}(\mathcal{M}(g); \alpha, \beta)| \leq c'' MN/p$ . This can be rewritten as

$$(1 - \delta) \frac{MN}{p} \leq \text{card}(\mathcal{M}(g) \cap [\alpha, \beta]) \leq (1 + \delta) \frac{MN}{p},$$

which proves the corollary.

#### 4. A FEW OPEN PROBLEMS

There are different points of view and ways to study the distribution of the elements of a certain sequence. But going further along the lines followed above, let us first notice that Theorem 1 and Corollary 1 applies not only to  $\mathcal{M}(g, \mathcal{J})$ , but for sets featuring certain patterns such as those generated when  $\mathcal{J}$  is replaced by unions of arithmetic progressions, also. This is easy to see, since

$$\mathcal{D}(\mathcal{M}_1 \cup \mathcal{M}_2) = \mathcal{D}(\mathcal{M}_1) + \mathcal{D}(\mathcal{M}_2),$$

for any sets  $\mathcal{M}_1, \mathcal{M}_2 \subset [0, 1]$  with  $\mathcal{M}_1 \cap \mathcal{M}_2 = \emptyset$ .

A further step in the evaluation of changes produced by the discrete logarithm function would be to evaluate the discrepancy when the original set ( $\mathcal{J}$  in the notation from the introduction) is additionally changed by a non linear transform. Such an example would require to estimate a sum such as, for instance,

$$\sum_{x \in \mathcal{J}} e_{p-1}(P(\log_g x)),$$

where  $P(x) = a_0 + a_1 x + \dots + a_n x^n$ , with  $a_0, \dots, a_n$  integers,  $a_n \not\equiv 0 \pmod{p}$  and  $n \geq 2$ .

Another spreading factor appears if more than one primitive root are involved. Let  $g_1, \dots, g_r$  be primitive roots mod  $p$  and let  $a, b_1, \dots, b_r$  be integers. Then the problem is to find a nontrivial estimate for the sum

$$\sum_{x \in \mathcal{J}} e_{p-1}(ax + b_1 \log_{g_1} x + b_2 \log_{g_2} x + \dots + b_r \log_{g_r} x).$$

Related to these questions is the problem that asks to study the changes produced by the discrete logarithm function in the order of its arguments. If the elements of  $\mathcal{L}(g, \mathcal{J})$  were randomly distributed in  $[0, p-1]$ , then comparing the size, for  $x_1, x_2 \in \mathcal{J}$  with  $x_1 < x_2$ , one expects that about half of the time  $\log_g x_1 < \log_g x_2$  and half of the time  $\log_g x_1 > \log_g x_2$ .

And similarly, for any fixed  $r \geq 2$ , when  $p \rightarrow \infty$ , all the  $r!$  possible arrangements among the numbers  $\log_g x_1, \dots, \log_g x_r \in [0, p-1]$  should occur with about the same frequency when  $(x_1, \dots, x_r)$  runs over  $\mathcal{J}^r$ .

## REFERENCES

- [1] W. Diffie and M. Hellman, *New directions in cryptography*, IEEE Trans. Inform. Theory **22** (1976), 644–654.
- [2] T. ElGamal, *A public key cryptosystem and a signature scheme based on discrete logarithms*, IEEE Trans. Inform. Theory **31** (1985), 469–472.
- [3] D. W. Kravitz, *Digital signature algorithm*, U.S. patent #5,231,668, 27 July, 1993.
- [4] A. Menzes, P. C. Van Oorschot and S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996, xxxvi+780 pp.
- [5] H. L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS, **84**. Providence, RI, 1994, xiv+220 pp.
- [6] A. Odlyzko, *Discrete Logarithms: The Past and the Future*, Preprint 1999.
- [7] W. D. Banks and I. E Shparlinski, *Exponential sums with polynomial values of the discrete logarithm*, Uniform Distribution Theory, **2** (2007), no. **2**, 67-72.

CRISTIAN COBELI, INSTITUTE OF MATHEMATICS OF THE ROMANIAN ACADEMY, P. O. BOX 1-764, BUCHAREST 70700, ROMANIA.

*E-mail address:* cristian.cobeli@imar.ro